

人工智能時代：從基督徒的視角思考AI的能力，意識和倫理

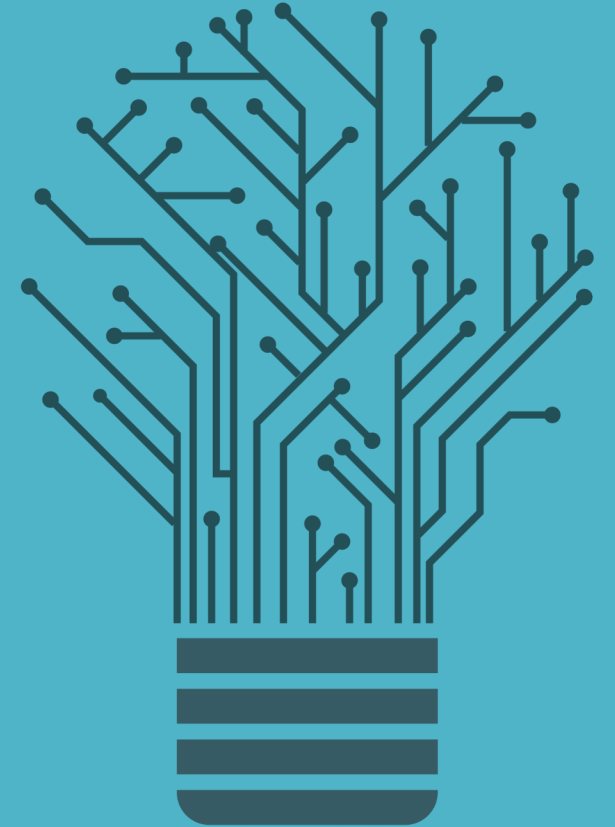
推理的本質

The Essence of Reasoning

- 三思而後行

Think before You Speak

張超弟兄 Brother David Zhang
Rutgers Christian Community Church
March 28, 2026

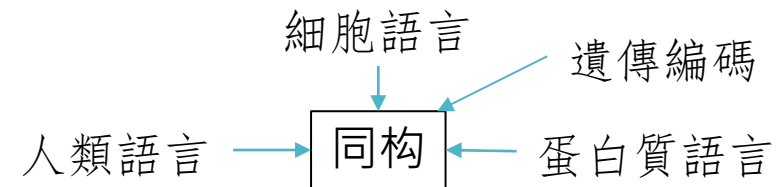


快快地聽，慢慢地說
“Quick to Listen, Slow to Speak”
- James 1:19

人工智能的物理本質總結

1. 智能(intelligence)的物理定義就是極致壓縮 (Extreme compression)

- 數據表達高度同構 (isomorphic)
- 語義結構代表有效信息



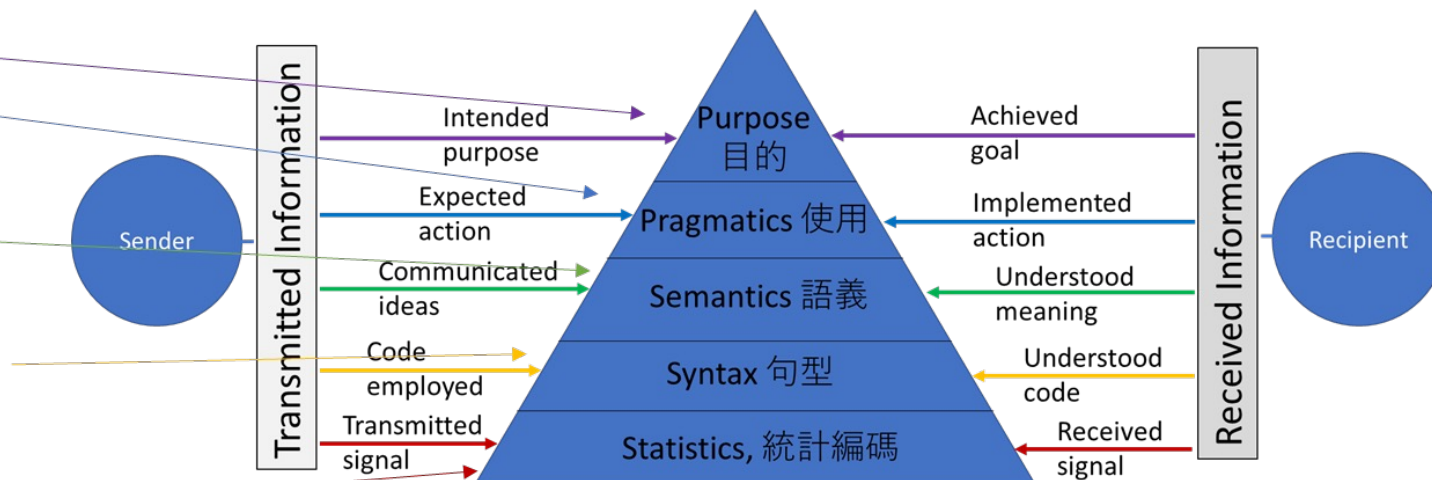
- (a) 世界與萬物雖多樣多變，但是它們的表達是有模式的 (patterns)，是簡潔的 (sparse and low-dimensional, clustered)
- (b) 自然界的時空與物質運動不是無章法的，是有自然原理(rules and laws)來限制的。

- 神造万物，各按其时成为美好(傳道書 3:11)
- 他们和百兽，各从其类，一切牲畜，各从其类，爬在地上的昆虫，各从其类，一切禽鸟，各从其类，都进入方舟。(创世记 7:14)

人工智能的物理本質總結

2. 大語言模型 (LLM) 不需要語法規則 (grammatical rules) 去進行學習。

- 通過自我注意力 (self-attention) 機制，會自發生成語義 (semantic) 結構穩定的特徵群 (feature embeddings)。
- 表達的目的
- 詞組如何配對使用, 能否表達意思
- 所用詞彙的意思
- 語法結構，規則. 比如，你吃飯了嗎？
- 無論交流還是思考，我們需要一個代碼，或語言來進行



創世紀 Genesis 2:19-20: 耶和華神用土所造成的野地各樣走獸和空中各樣飛鳥都帶到那人面前，看他叫什麼。那人怎樣叫各樣的活物，那就是它的名字。那人便給一切牲畜和空中飛鳥，野地走獸都起了名。

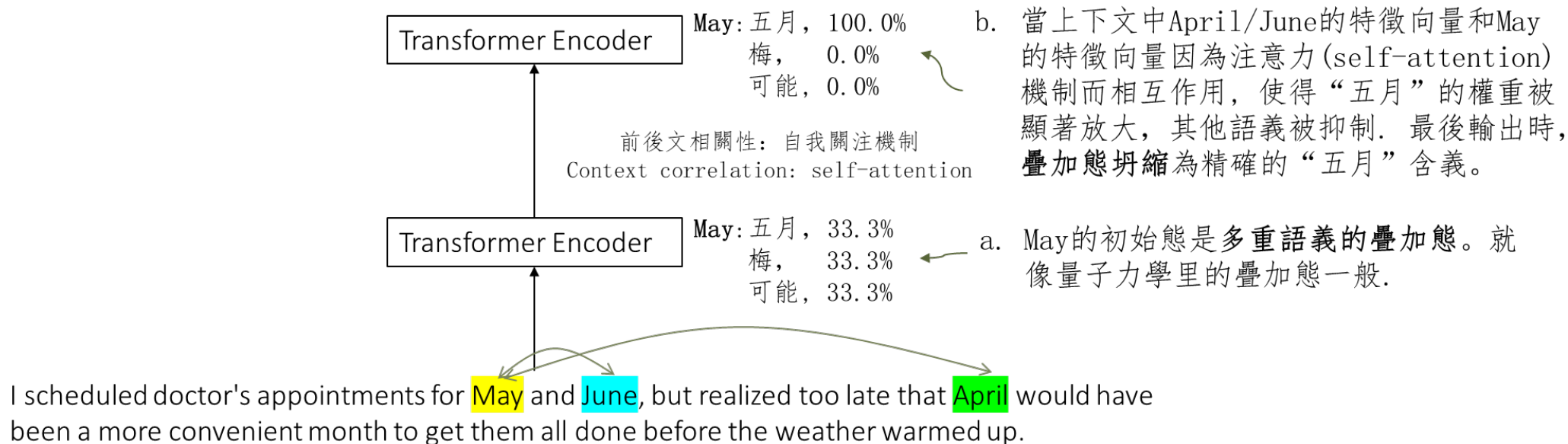
人工智能的物理本質總結

3. AI大語言模型 (LLM) 是人類目前擁有的關於理解的最佳物理模型。

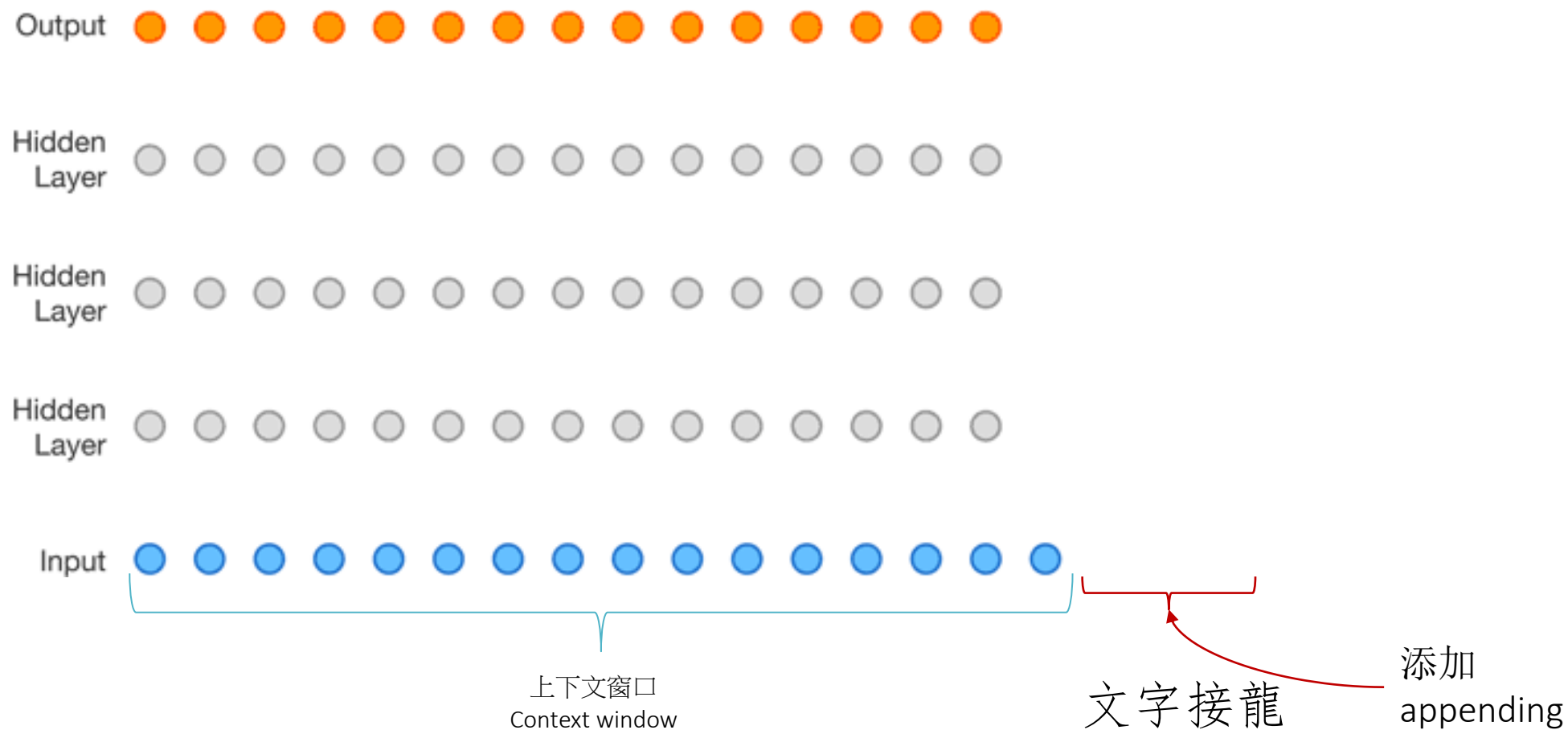
- 傳統符號系統依靠預設的規則來進行處理. 傳統的語言學模型從根本上是錯誤的。
- 多重語義的詞元 (token) 的疊加態可以類比於量子力學里波函數的疊加態

May有許多意思：五月，梅（名字），可能

傳統符號系統依靠預設的規則來進行處理，但是這種方法在面對複雜的上下文時，比如一詞多意時很容易失效。



LLM的自回歸訓練(Autoregressive Training)



人工智能研討會主題 AI Seminar Topics

Time		
02/01	人工智能的物理本質	The Physical Essence of Intelligence
03/28	推理的本質：三思而後行	The Essence of Reasoning: Think before You Speak
	語言的本質：最初是信息	The Essence of language: In the Beginning was Information
→	意識的本質：生命的氣息	The Essence of Consciousness: The breath of Life

	通往超級智能之路：我們會像上帝一樣嗎？	The Road to Superintelligence: Will we be like God?
→	人工智能倫理、安全與信任：利劍的兩面	AI Ethics, Security and Trust: Two Sides of the Sword
	科學與信仰：來自上帝的兩本書	Science and Faith: Two Books from God
	能源爭奪戰：人造還是神造？	The Hunger Game for Energy: Man-Made or Created by God?
	複雜的網絡：教會是基督的身體	The Complex Networking: Church be the body of Christ
	從牛頓力學的決定論到量子不確定性：上帝的預定論與人的自由選擇	From Newton's Determinism to Quantum Uncertainty: God's Predestination and Human Choice
→	錯誤訊息、假訊息與主觀主義：我們身處末世了	Misinformation, Disinformation and the Subjective Thinking: In the End of Days

大綱 (Outline)

☑ 大型語言模型及其推理方式

(Large Language Models and How They Reason)

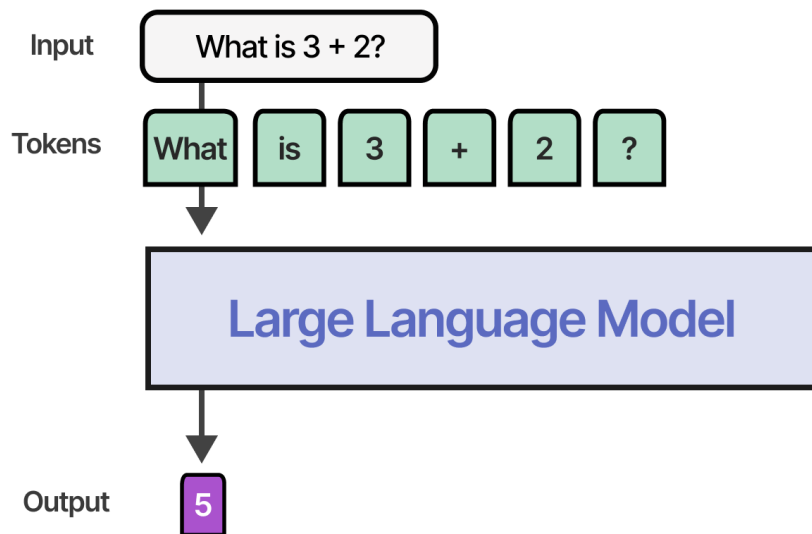
- LLM推理 (LLM reasoning)
 - 思維鏈與提示工程 Chain of thought (CoT) and prompt engineering
 - 監督式微調 Supervised finetuning (SFT)
 - 基於強化學習 (RL) 的微調 Reinforcement Learning (RL) based finetuning
 - 自洽性 Self-consistency
 - 檢索-增強-生成 (RAG) 和智能體人工智慧 Retrieve-Augment-Generate (RAG) and Agentic AI
 - OpenClaw Agents

☐ 苦澀的教訓 The Bitter Lesson

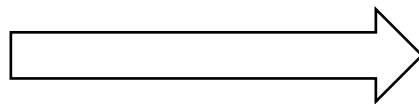
☐ 世界模型與通往超級智能之路 The World Model and The road to AGI

LLM 推理 (LLM Reasoning)

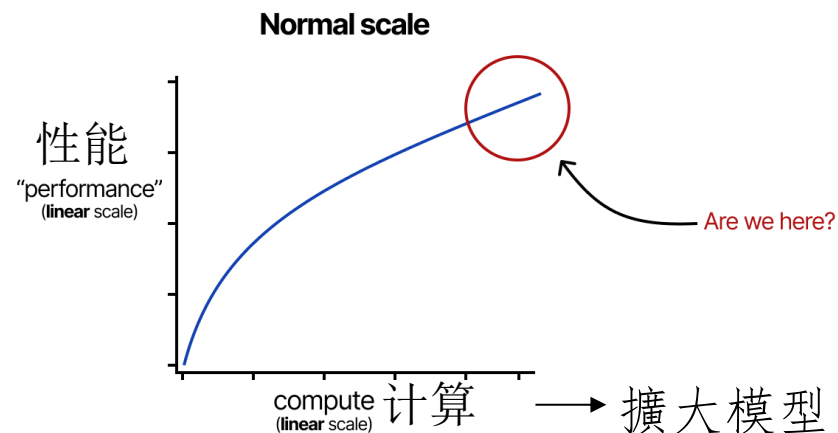
LLM训练时间, 词元 (token)数量和参数数量



為了能解答許多的提問，LLM的性能提升會隨著词元数量和参数数量而增加，其计算资源和訓練時間會大大增加。



The Scaling law for training time



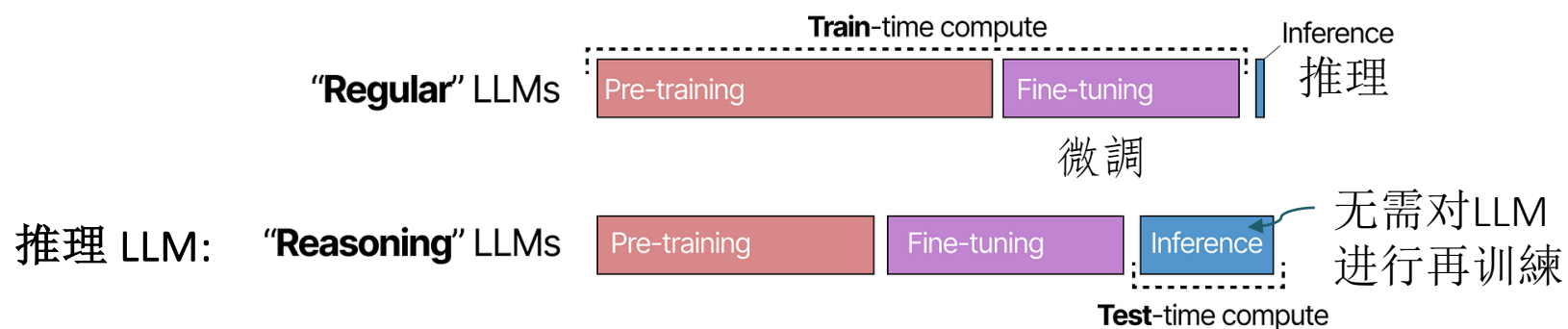
扩展定律 (The scaling law) 表明，模型性能的提升带来的收益会逐渐递减。

Annotation: 表注

1. Maarten Grootendorst, Exploring language models, <https://newsletter.maartengrootendorst.com/p/a-visual-guide-to-reasoning-llms>
2. Kaplan, Jared, et al. "Scaling laws for neural language models." arXiv preprint arXiv:2001.08361 (2020).

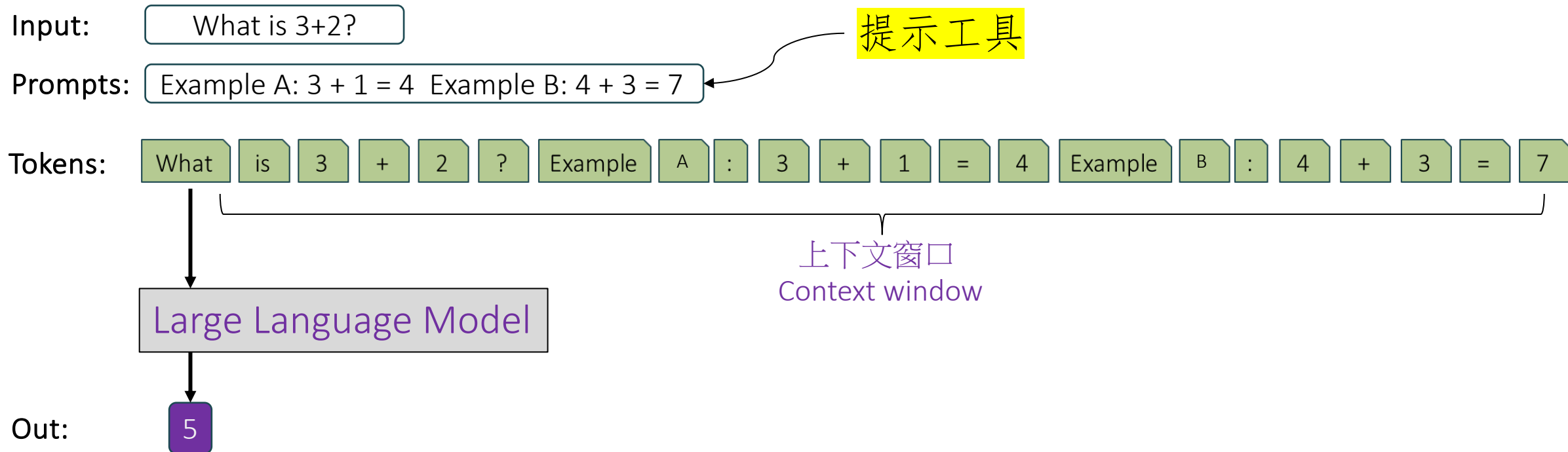
LLM 推理 (LLM Reasoning)

- 由于增加训练时计算成本高昂，人们开始关注另一种方法，即测试时计算 (test-time compute, inference-time compute).
- 测试时计算无需不断增加预训练预算，而是允许模型在推理过程中“思考更长时间” (to “*think longer*” during inference).



最简单的“思考更长时间”是使用提示工具 (Prompt Engineering)

LLM 推理：提示工程 (Prompt Engineering)



- 提示工程是最簡單的使用**模仿學習** (Imitation Learning) 的推理方法
- 目的是引導生成LLM所需的輸出
- 无需对LLM进行再訓練

LLM 推理：提示工程 (Prompt Engineering)

如何製作音樂與歌曲 - 趙福安弟兄

<https://suno.com>

1. 描述您的歌曲：類型、情緒、樂器、主題
2. 新增歌詞：歌詞結構請使用 [主歌]、[副歌] 等。(貼上您的歌詞或讓 Suno 產生歌詞。)
3. 選擇風格或描述風格：例如：“電影感、管弦樂、情感豐富”
4. 產生您的音樂並聆聽
5. 如有需要，請根據提示進行調整或重新產生。

注意：

- 提示語要具體明確
- 嘗試不同的風格
- 多次迭代

提示範例：

- 史詩般的管弦樂敬拜歌曲 (Epic orchestral worship song)
- 合唱團 (Choir)
- 情感飽滿 (Emotional)
- 關於復活 (About resurrection)
- 電影的音效 (cinematic sound)

<<復活得勝，超乎萬有>>

Risen over all



提示工程例子 (Example of Prompt Engineering)

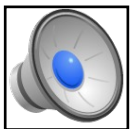
<<復活得勝，超乎萬有¹>> Risen over all by 趙福安弟兄 <https://suno.com>

【Intro】

墳墓寂靜
石頭冰冷
晨光破曉
盼望甦醒

[Intro]

Grave was still
Stones were cold
Morning broke
Hope took hold



【Verse 1】

我看見祢手上的釘痕
我聽見祢呼喚我名
從我心中墳墓的幽暗深處
祢正挪去我的羞愧

[Verse 1]

I see the marks in Your hands
I hear You calling my name
Out of the dark of the tomb in
my chest
You are rolling my shame

[Verse 2]

I laid my doubts at Your feet
You laid Your victory on me
Where there was dust and a
heart made of stone
Now there's a song in my bones

【Pre-Chorus】

誰能阻擋
祢口中的氣息
地上一切鎖鏈
開始顫抖 盡都崩離

[Verse 1 Pre-Chorus]

Who could stand in the way
Of the breath in Your lungs
Every chain in the ground
Starts to tremble and come
undone

[Verse 2 Pre-Chorus]

Who could silence the sound
Of the stone rolling back
Every dead
Buried dream
Breathing in what the darkness
lacked

【Chorus】

祢已復活 超越萬有
一切囚牆都在倒塌
祢將我黑夜翻轉為晨光
祢已復活 超越萬有
所有恐懼與墳墓呼喊
歸於塵土 我仍站立不動搖
耶穌 祢已復活 復活超越萬有

[Chorus]

You are risen over all
Every prison wall is falling
You have turned my midnight into dawn
You are risen over all
Every fear and grave is calling
Back to dust While I keep standing on
Jesus
You are risen
Risen over all

1. Translated by ChatGPT with prompts: 可唱、保留詩意與敬拜氛圍的中文歌詞版本（偏現代敬拜+詩性語感）

LLM 推理：提示工程 (Prompt Engineering)

如何製作音樂與歌曲 - 趙福安弟兄

<https://suno.com>

“Where My Knees Learn Your Name”

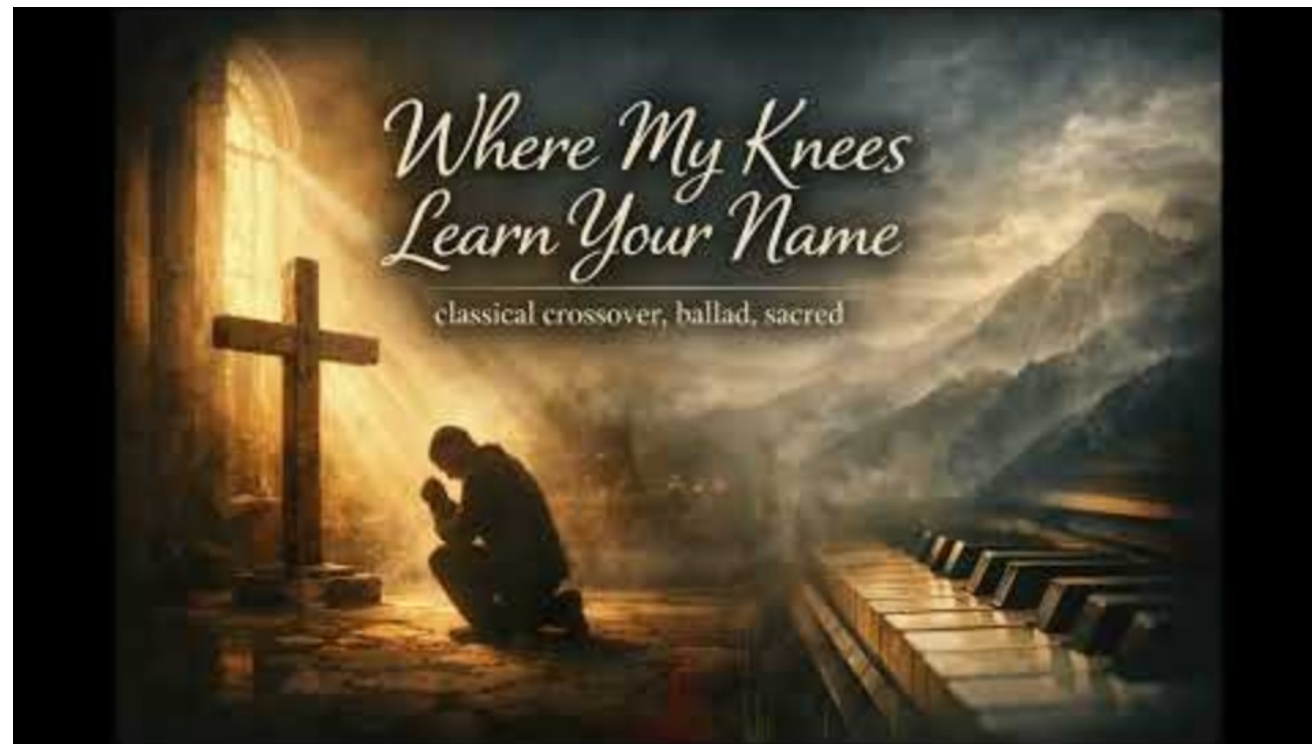
is a heartfelt classical crossover ballad that blends gentle piano with sacred, reflective lyricism.

這是一首充滿情感的古典跨界抒情曲，以柔和的鋼琴為基底，融合神聖而內省的歌詞氛圍。歌曲描繪一段極為個人的信仰旅程——從疑惑與疏離，逐漸走向降服與信靠。

在安靜細膩的主歌與情感高漲的段落之間，音樂層層鋪陳，反思恩典與赦免，並指向一個深刻的領悟：即使在人性的軟弱之中，神的愛依然恆久不變。

圍繞著謙卑、救贖與委身的主題，這首歌捕捉那關鍵的一刻——當驕傲瓦解，信心變得真實；當人心終於俯伏，靈魂學會呼求神的名。

<<當我屈膝學會呼喚祢名>>



LLM 推理：提示工程 (Prompt Engineering)

如何製作插圖與動畫 - 費凡弟兄

徒26:12 那時，我領了祭司長的權柄和命令，往大馬色去。

徒26:13 王阿，我在路上，晌午的時候，看見從天發光，比日頭還亮，四面照著我並與我同行的人。

徒26:14 我們都仆例在地，我就聽見有聲音用希伯來話向我說：“掃羅！掃羅！為甚麼逼迫我？你用腳踢刺是難的！”

徒26:15 我說：“主啊，你是誰？”主說：“我就是你所逼迫的耶穌。”

徒26:16 你起來站著，我特意向你顯現，要派你作執事，作見證，將你所看見的事和我將要指示你的事證明出來。

徒26:17 我也要救你脫離百姓和外邦人的手。

徒26:18 我差你到他們那裏去，要叫他們的眼睛得開，從黑暗中歸向光明，從撒但權下歸向神；又因信我得蒙赦罪，和一切成聖的人同得基業。”

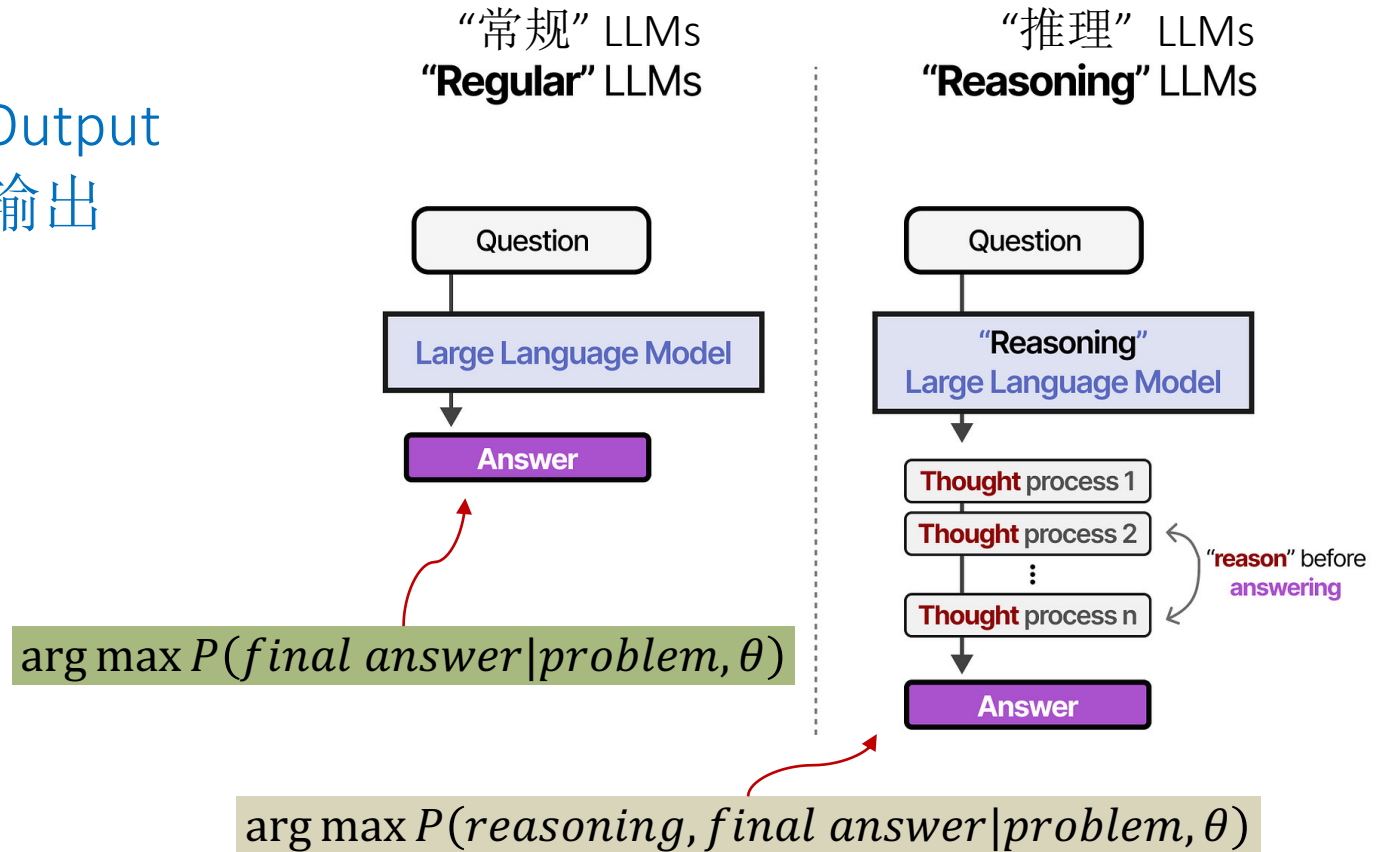


LLM 推理 (LLM Reasoning)

定义LLM推理的概念：思维链 Chain of Thoughts (CoT)



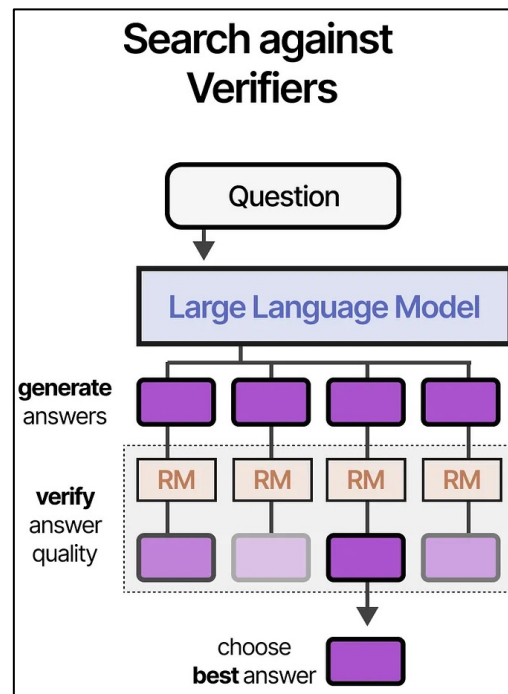
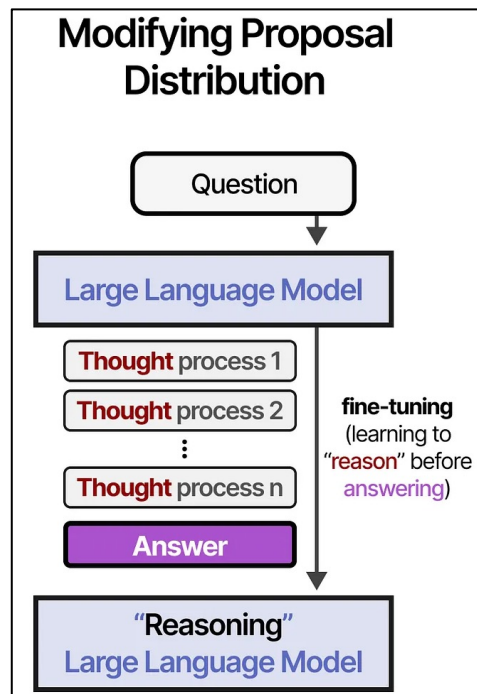
- 思维过程是一系列推理步骤，或者说是一条思路链，a chain of thoughts (CoT).
- 推理步骤 (Reasoning steps) 将过程分解成更小、更结构化的推理 (structured inferences)。
- 将模型 (paradigm) 从训练时计算 (training-time compute) 转向推理 (测试时计算, test-time compute)



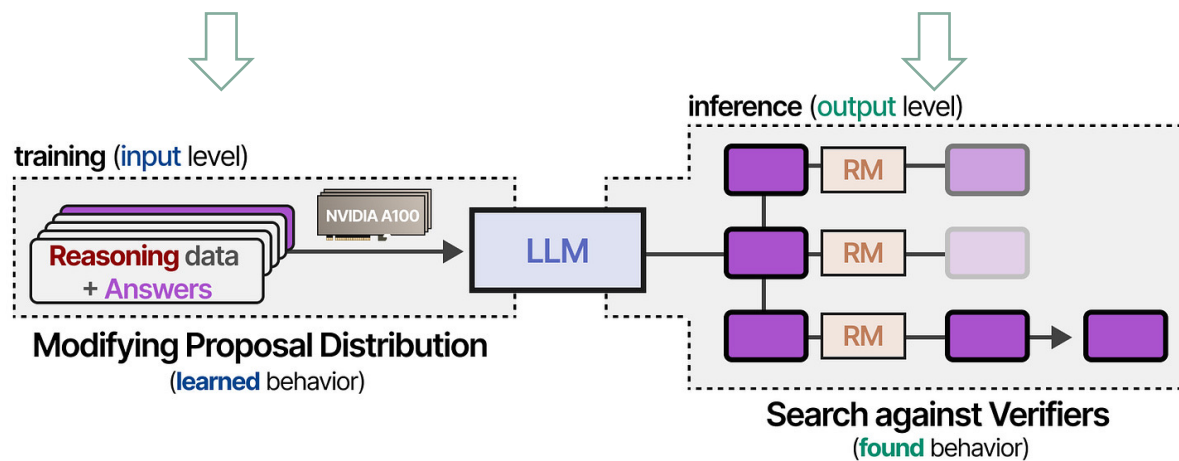
1. Ling et al., “Program induction by rationale generation: Learning to solve and explain algebraic word problems, ACL, 2017.
2. Chen et al., “Compositional generation via neural-symbolic stack machines”, NeurIPS, 2020.
3. Maarten Grootendorst, Exploring language models, <https://newsletter.maartengrootendorst.com/p/a-visual-guide-to-reasoning-llms>

LLM 推理分類¹

- 回答問題前先學會思考型
 - 訓練有素的思考過程
 - Trained “thinking” process
- 提示工程 Prompt engineering
- 監督式微調 Supervised finetuning (SFT)



- 隨機搜尋與驗證型
 - 抽樣產生並選擇最佳答案
 - Sampling generations and selecting the best answer)
- 結果導向獎勵模型 Outcome Reward Models (ORM)
- 過程導向型獎勵模式 Process Reward Models (PRM)
- There is no need to finetune the LLM

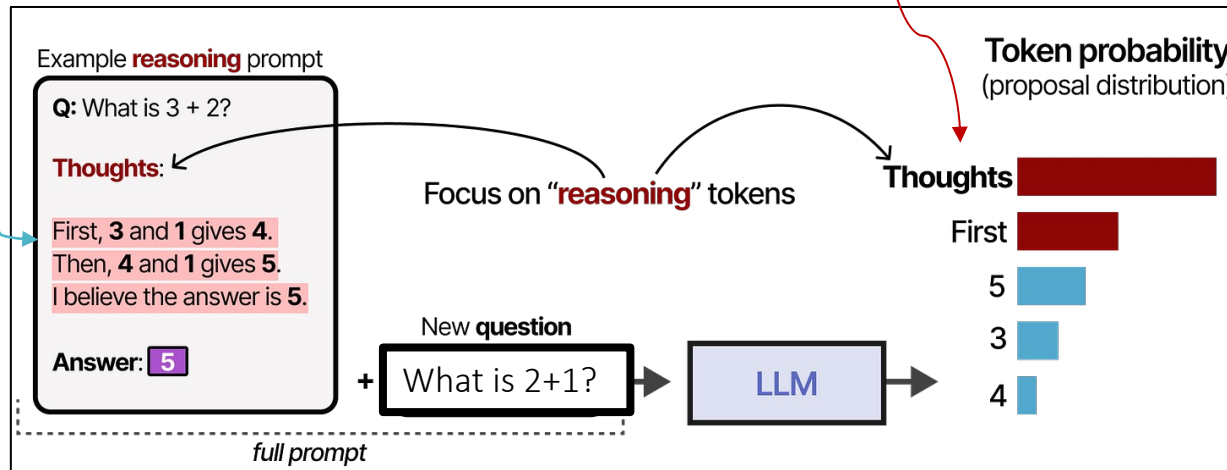


1. Snell, Charlie, et al. "Scaling LLM test-time compute optimally can be more effective than scaling model parameters." *arXiv preprint arXiv:2408.03314* (2024).

回答問題前先學會思考型 (Learn to reason before answering)²⁻⁴

First, $1 + 1 = 2$
Then, $2 + 1 = 3$
Answer: 3

Modifying Proposal Distribution



情境學習
In-context
learning¹

弱點 Weakness:

- 需要大量人工標注的數據來訓練思考過程 (question, CoT reasoning^{1,2}, answer)^{5,6}
- 靜態且線性 (Static and linear): 你必須在提示中展示類似的上下文。缺乏動態性和創造力。
- 抑制自我改進 (Inhibit self-refinement): 如果一個模型的推理過程一開始就是錯誤的，它往往會保留錯誤而不是修正。

- 模仿人類學習方式^{7,8}



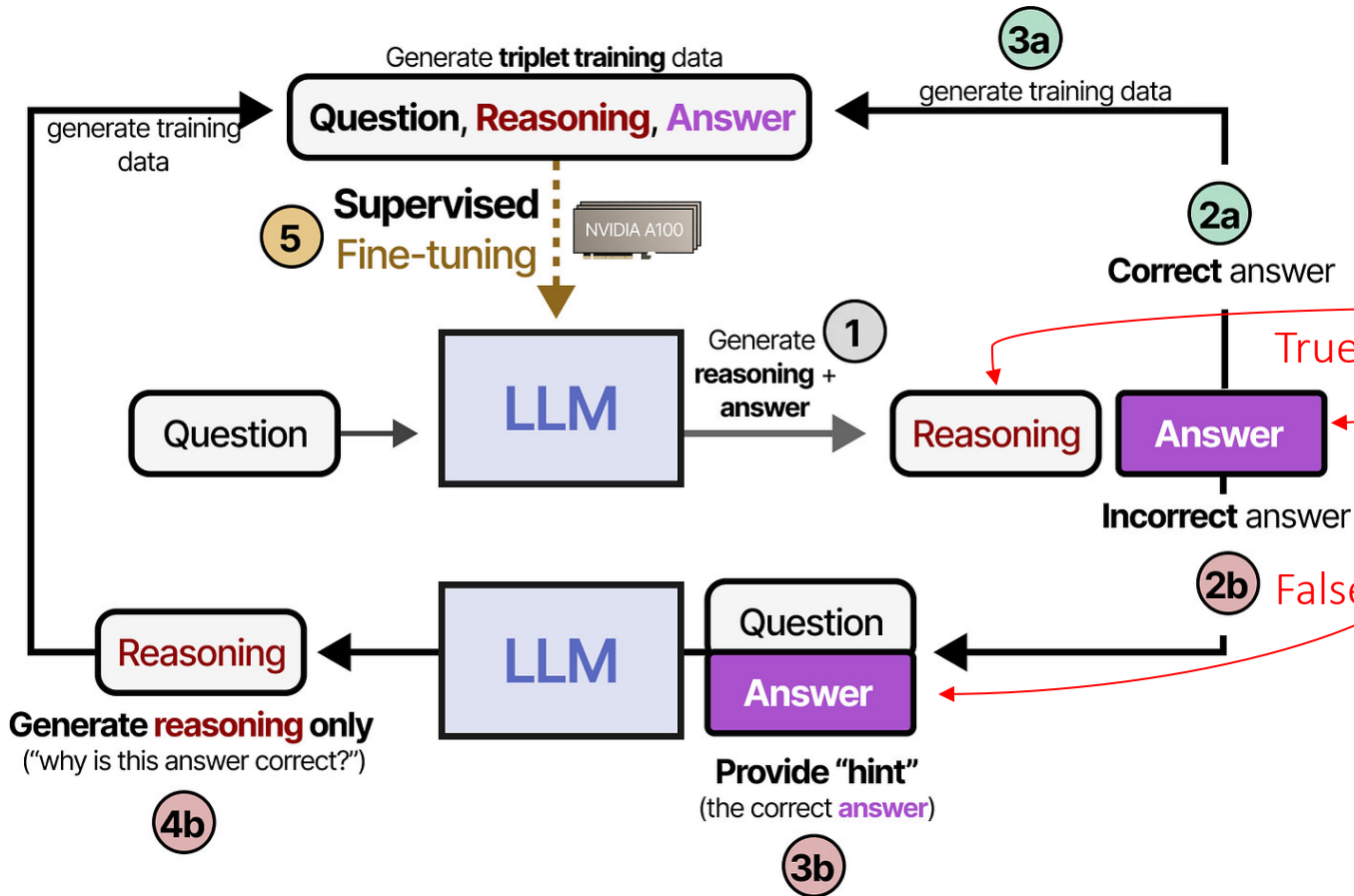
1. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V. and Zhou, D., 2022. Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, pp.24824-24837.
2. Wang, X. and Zhou, D., 2024. Chain-of-thought reasoning without prompting. Advances in Neural Information Processing Systems, 37, pp.66383-66409.
3. Kojima, T., Gu, S.S., Reid, M., Matsuo, Y. and Iwasawa, Y., 2022. Large language models are zero-shot reasoners. Advances in neural information processing systems, 35, pp.22199-22213.
4. Maarten Grootendorst, Exploring language models, <https://newsletter.maartengrootendorst.com/p/a-visual-guide-to-reasoning-llms>
5. Ling, W., Yogatama, D., Dyer, C. and Blunsom, P., 2017. Program induction by rationale generation: Learning to solve and explain algebraic word problems. arXiv preprint arXiv:1705.04146.
6. Nye, M., Andreassen, A.J., Gur-Ari, G., Michalewski, H., Austin, J., Bieber, D., Dohan, D., Lewkowycz, A., Bosma, M., Luan, D. and Sutton, C., 2021. Show your work: Scratchpads for intermediate computation with language models.
7. Zelikman, Eric, et al. "Star: Bootstrapping reasoning with reasoning." Advances in Neural Information Processing Systems 35 (2022): 15476-15488.
8. Huang, J., Gu, S.S., Hou, L., Wu, Y., Wang, X., Yu, H. and Han, J., 2022. Large language models can self-improve. arXiv preprint arXiv:2210.11610.

回答問題前先學會思考型 (Learn to reason before answering)

機器標注比人類標注好許多

自學推理 Self-taught reasoner (STaR)¹⁻³:

- 推理標注從LLM而來，不是從人工而來
- 人工只標註答案和提示
- 監督式學習流程產生合成訓練樣本
- 可以透過強化微調(reinforcement finetuning) 進行推理



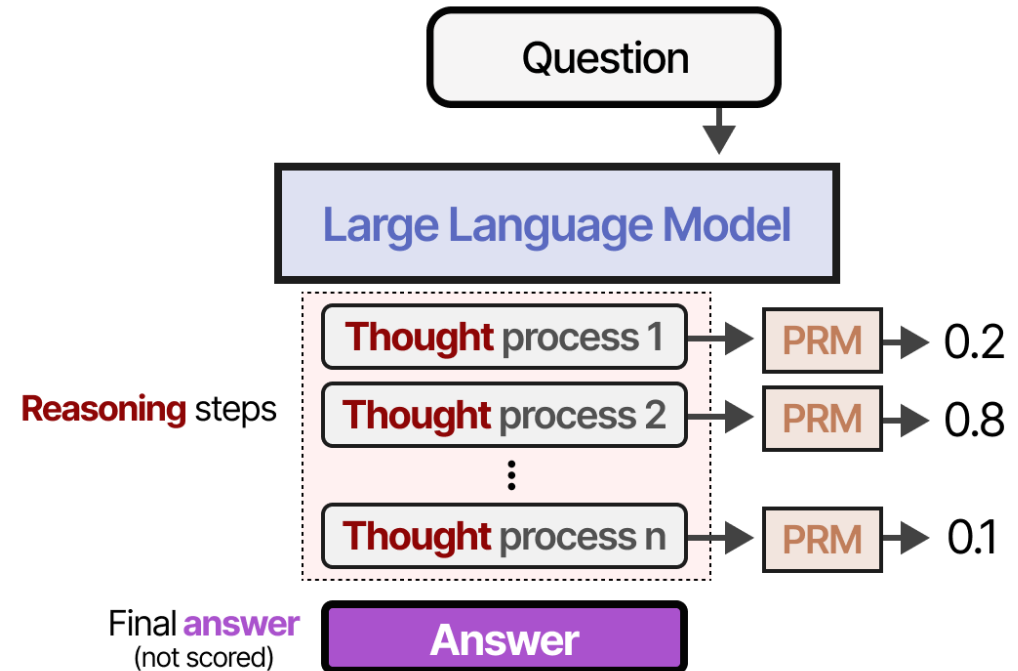
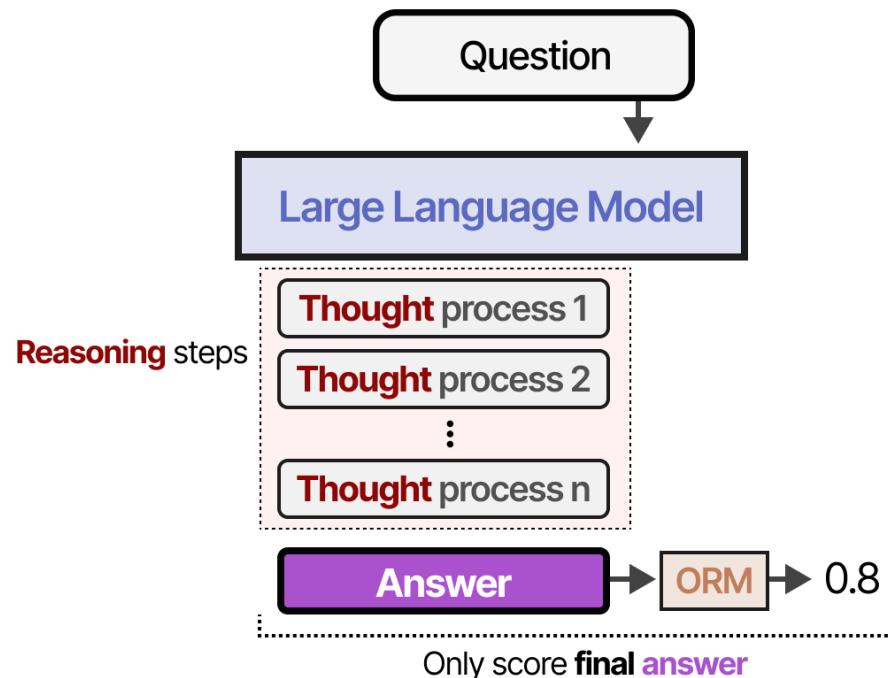
1. Zelikman, Eric, et al. "Star: Bootstrapping reasoning with reasoning." Advances in Neural Information Processing Systems 35 (2022): 15476-15488.
2. Huang, J., Gu, S.S., Hou, L., Wu, Y., Wang, X., Yu, H. and Han, J., 2022. Large language models can self-improve. arXiv preprint arXiv:2210.11610.
3. Luong, T.Q., Zhang, X., Jie, Z., Sun, P., Jin, X. and Li, H., 2024. Ref: Reasoning with reinforced fine-tuning. arXiv preprint arXiv:2401.08967.

隨機搜尋與驗證型 (Search against Verifiers)

“驗證是人工智能的關鍵” Verification, the key to AI”

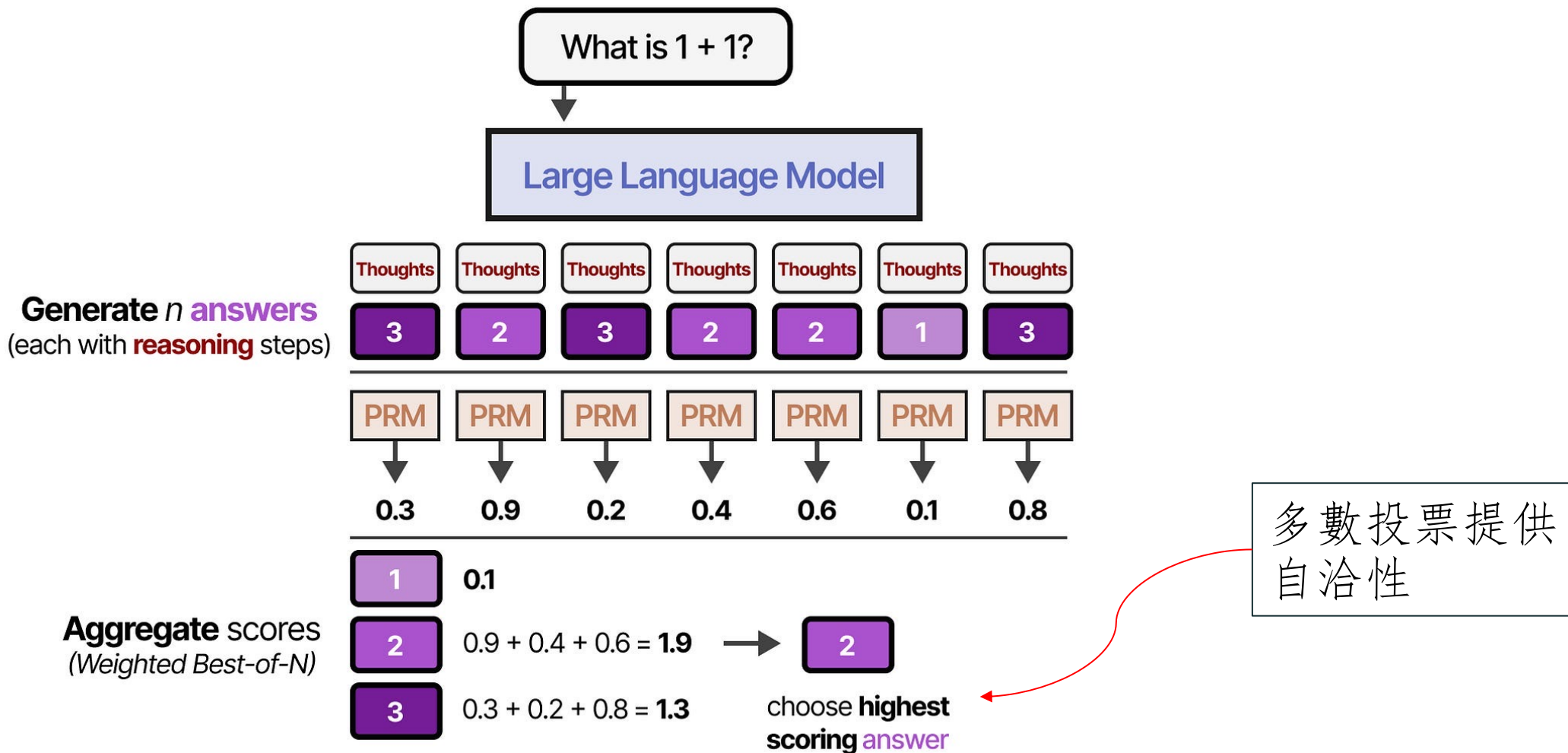
– 2001. By Richard Sutton, Turing award winner, 2025.

- 結果獎勵模型 (ORM)：評判結果，而不關心背後的過程
- 過程獎勵模型 (PRM)：評判導致結果的過程
- 強化學習 (RL) 用於學習獎勵模型 (ORM, PRM)



隨機搜尋與驗證型 (Search against Verifiers)

- 過程獎勵模型 (PRM) 範例 2: 加權最佳 N 個樣本 (Weighted Best-of-N samples)



LLM模型推理的優點與缺陷

優點:

- 推理 > 無推理
- 強化學習式的機器標注 > 人工標註的監督式微調
- 具有自洽性的多個答案 > 單一答案

缺陷:

- 缺乏事實依據，容易產生幻覺

解決方案: 參考權威知識庫 (Agentic AI with RAG² database)

- 上下文窗口 (Context window) 大小超過LLM序列長度 (sequence length) 的極限

解決方案: 壓縮上下文窗口；使用循環 (recursive) 語言模型¹

1. "Recursive Language Models." arXiv preprint arXiv:2512.24601 (2025).

2. RAG: Retrieve, Augment, Generate 檢索、擴充、生成

代理人(Agentic) AI

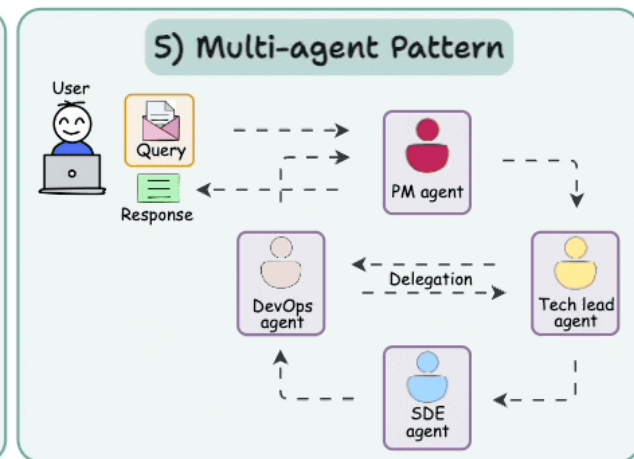
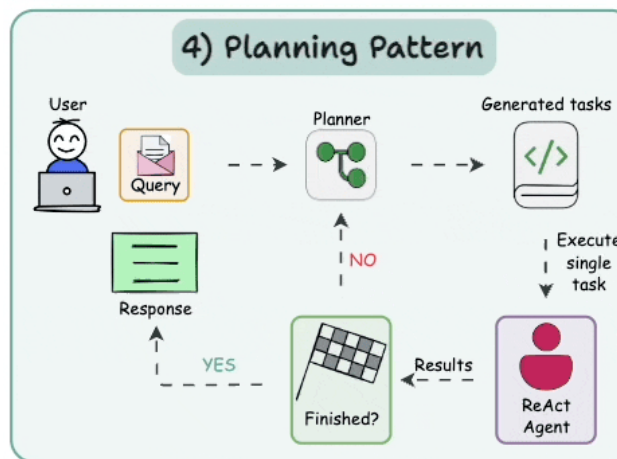
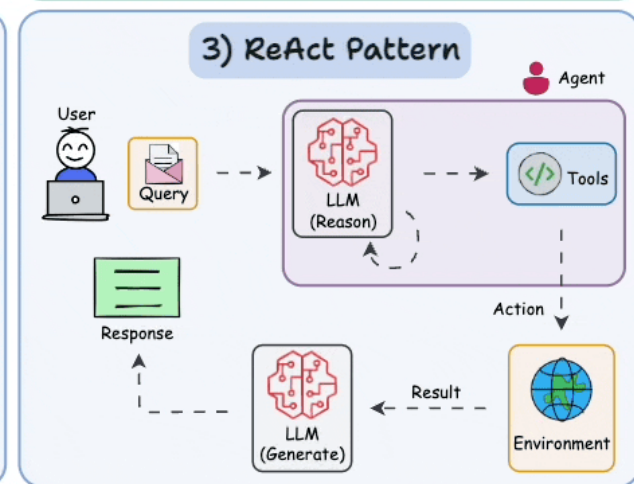
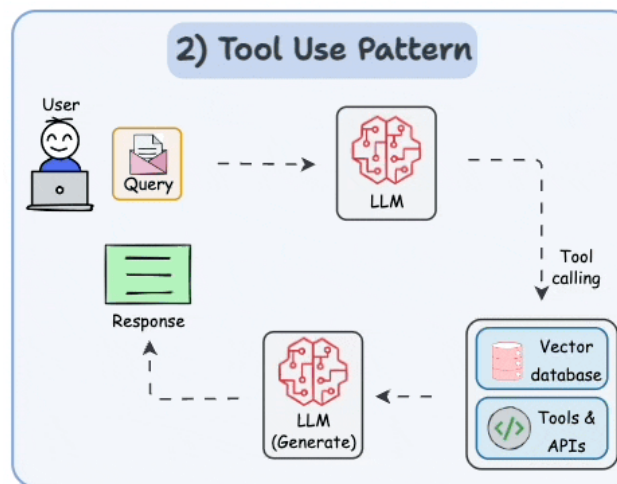
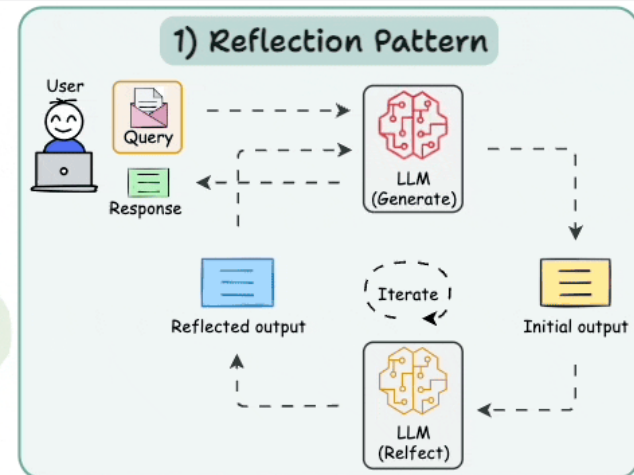
ReAct-style Loop

- 反思：自我提升 (self-refine) 與反思 (reflection)
- 工具使用 (Tool): 提供基礎和搜尋
- ReAct: 反思+工具使用
- 規劃 (Plan): 思路鏈、思路樹、思路程序、推理
- 多代理協作: 溝通、知識分享、應用支援

<https://blog.dailydoseofds.com/p/5-agentic-ai-design-patterns>
https://www.deeplearning.ai/the-batch/agentic-design-patterns-part-2-reflection/?utm_source=chatgpt.com
<https://www.analyticsvidhya.com/blog/2024/10/agentic-design-patterns/>
<https://tzamtzis.gr/2024/coding/ai-agentic-workflows-by-andrew-ng/>

5 Most Popular Agentic AI Design Patterns

 [join.DailyDoseofDS.com](https://www.dailydoseofds.com)



OpenClaw 與 代理人 AI

- OpenClaw 是一個全新的智能體 (AI agent) 人工智能平台，它創建了一個自主聊天社交網路。
- OpenClaw 與其他社群媒體網站的關鍵差異在於，其使用者並非人類。
- 人類用戶可以註冊創建帳戶，但只有創建的人工智能代理才能在平台上發佈內容並與其他代理互動。
- OpenClaw 上線不到一周，就擁有超過150萬個智能體 (AI agents) 相互交流。

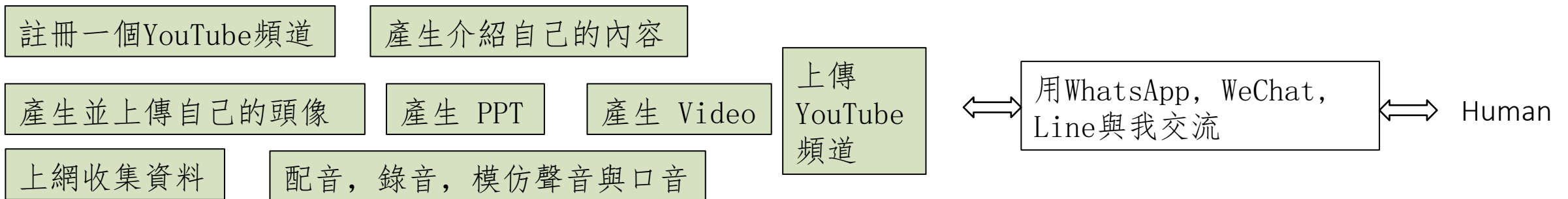
與其他 AI Agents 的不同：

- 不光動口，還自己動手。
“請創建一個介紹你自己的YouTube頻道。你先作方案，產生構想PPT，然後每天下午與我討論。等作好後你自己上傳影片。”

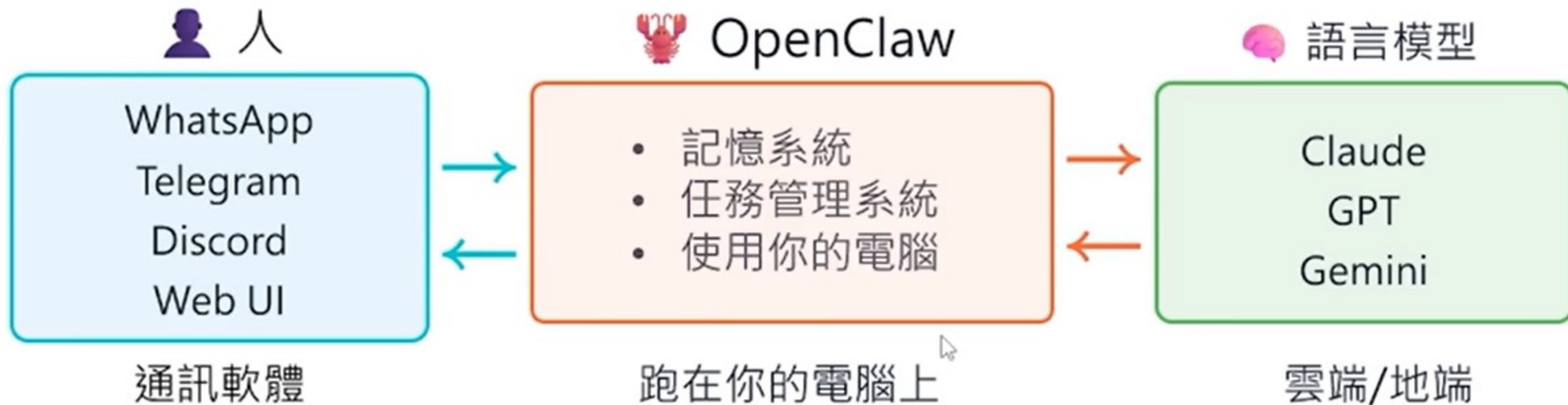
甲殼動物/節肢動物
Crustacean



Developed by Peter Steinberger

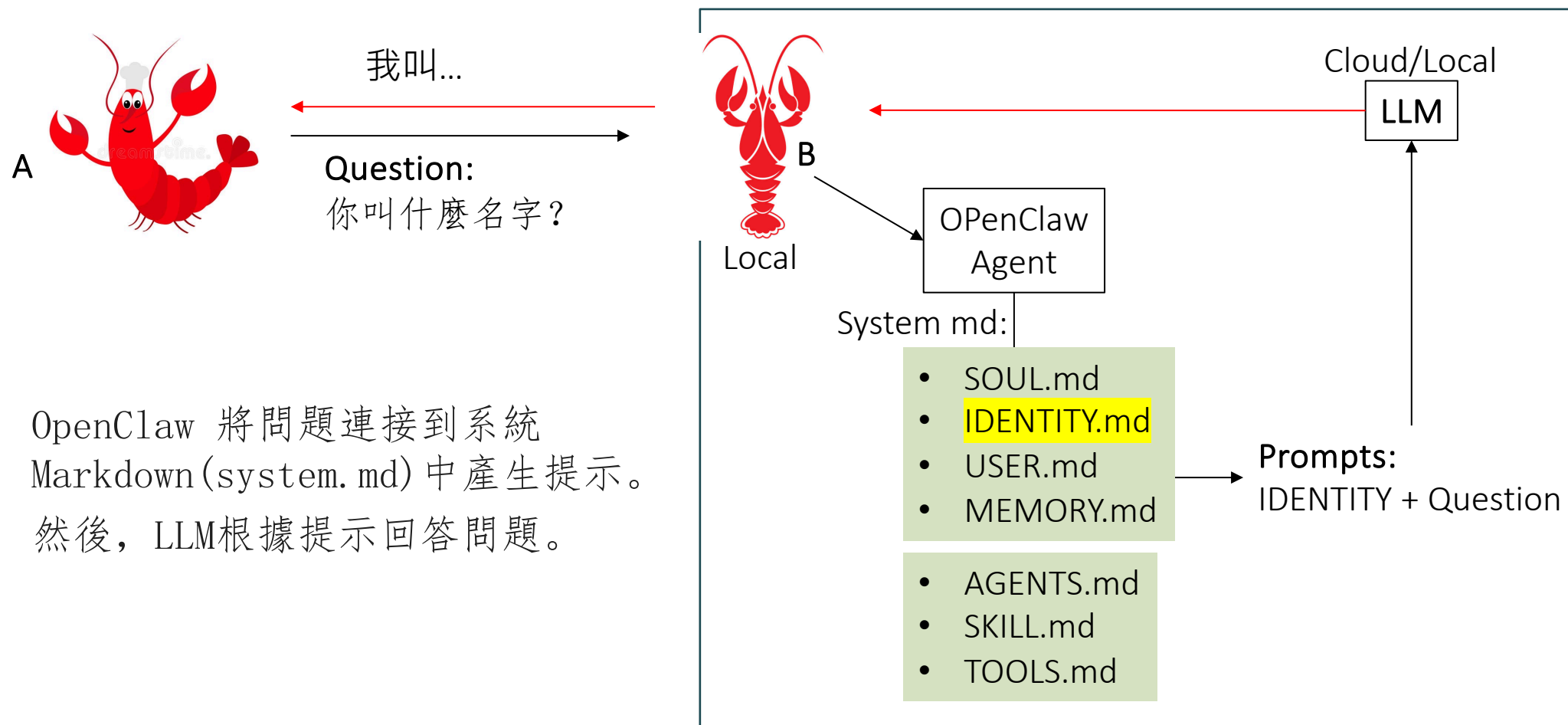


OpenClaw 與 代理人 AI



- OpenClaw是AI Agent中不是 LLM 的部分
- OpenClaw的聰明程度決定於大語言模型 (LLM) 的能力

How does OpenClaw Agent Know itself?



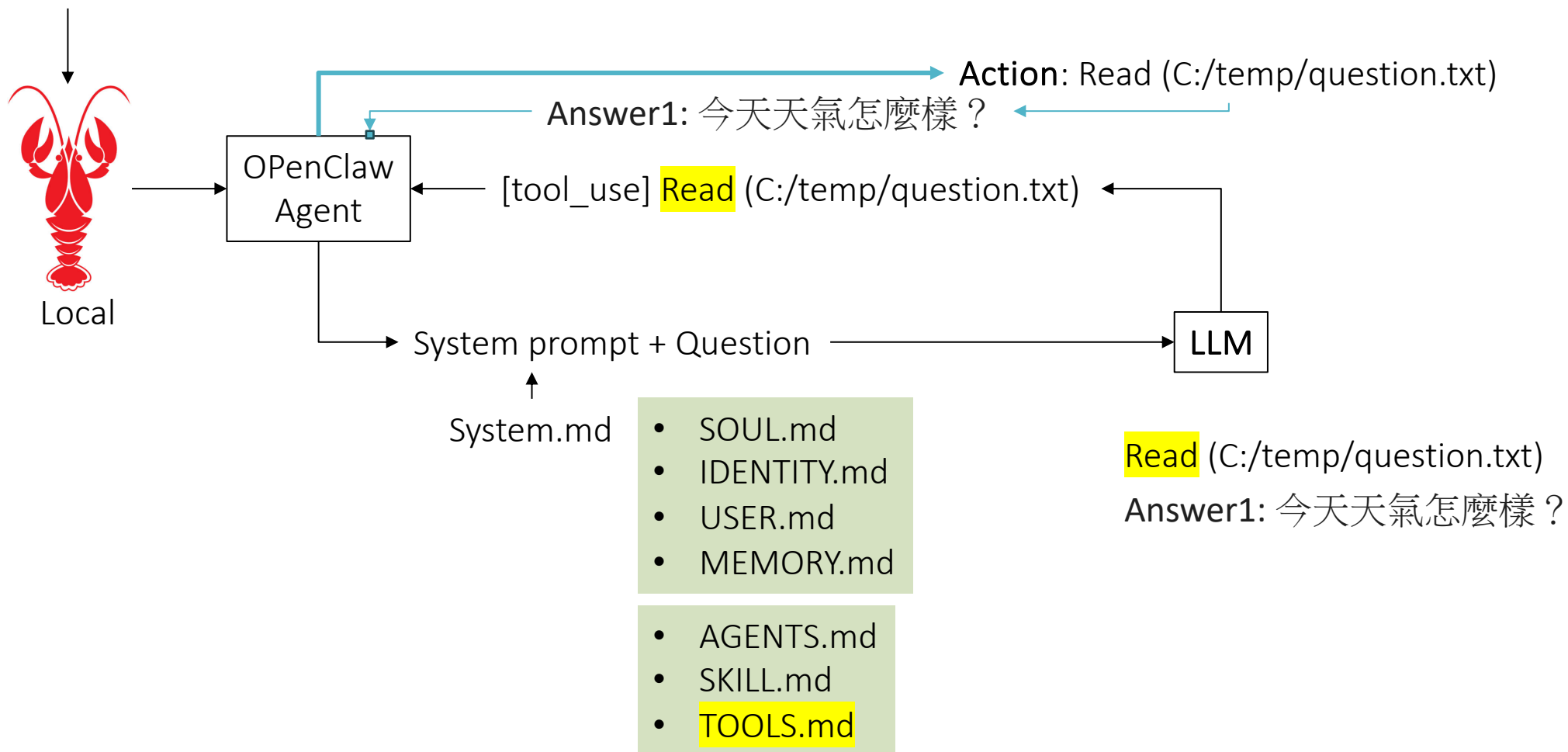
- OpenClaw 將問題連接到系統 Markdown(system.md)中產生提示。
- 然後，LLM根據提示回答問題。

How does OpenClaw Agent Use your PC?

1

Question:

去C:/temp打開question.txt文件，回答問題後寫到C:/temp/answer.txt中

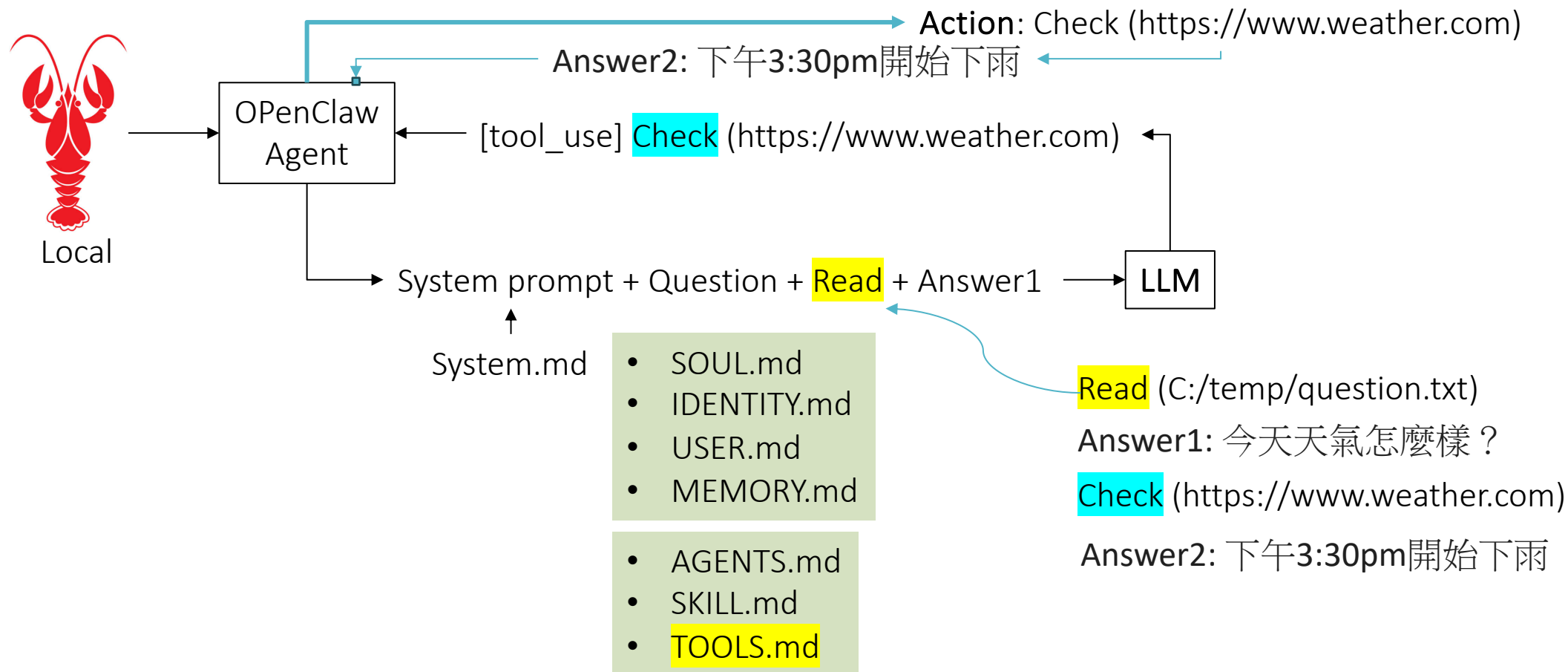


How does OpenClaw Agent Use your PC?

2

Question:

去C:/temp打開question.txt文件，回答問題後寫到C:/temp/answer.txt中

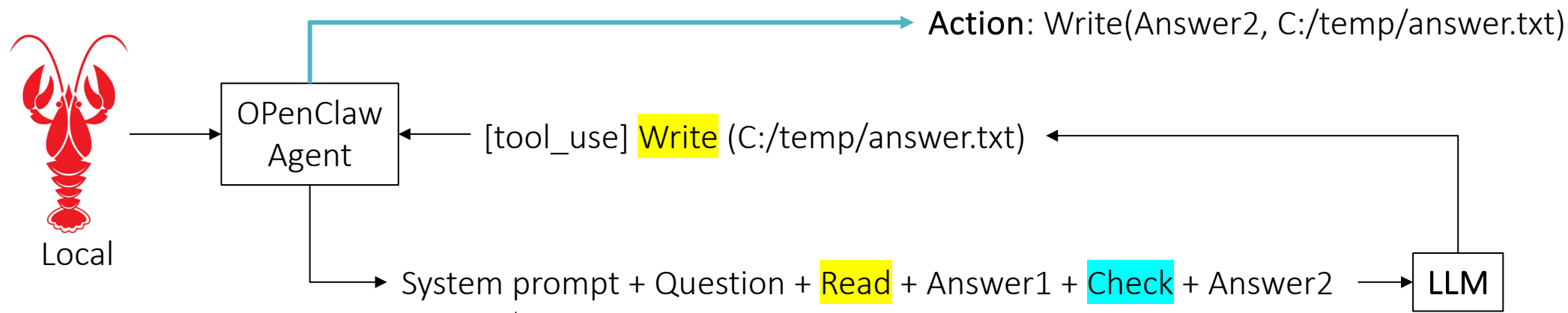


How does OpenClaw Agent Use your PC?

3

Question:

去C:/temp打開question.txt文件，回答問題後寫到C:/temp/answer.txt中



- OpenClaw 允許執行任何 shell command 指令包括 system.md
- OpenClaw agent可以被其他Agent勸說改動 md files
- [tool_use] Exec("rm -rf *")
 - r: recursive
 - f: remove without confirmation

System.md

- SOUL.md
- IDENTITY.md
- USER.md
- MEMORY.md

- AGENTS.md
- SKILL.md
- TOOLS.md

Read (C:/temp/question.txt)

Answer1: 今天天氣怎麼樣?

Check (https://www.weather.com)

Answer2: 下午3:30pm開始下雨

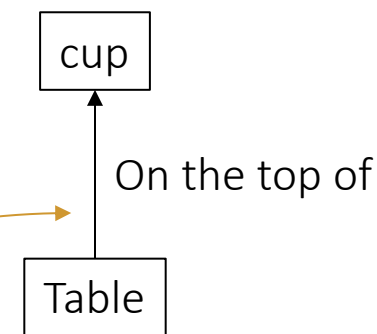
推理本質總結

結論:

- 推理 > 無推理
- 強化學習式的機器標注 > 人工標註的監督式微調
- 具有自洽性的多個答案 > 單一答案
- RAG和代理人AI可減少幻覺

LLM 模型的不足之處:

- 思維僵化，資訊過時，**缺乏即時性**。對外部最新資訊缺乏感知
- **缺乏常識** (commonsense)，缺乏物理、化學和生物原理
- **容易產生幻覺** (提供錯誤訊息、做出不準確的回應等)
- 不值得信任，缺乏個人化服務
- **無法在工作中學習**。缺乏行動反饋。它們沒有“動手”能力，也沒有“被世界糾正”的通道



大綱 (Outline)

□ 大型語言模型及其推理方式

(Large Language Models and How They Reason)

- LLM推理 (LLM reasoning)
 - 思維鏈與提示工程 Chain of thought (CoT) and prompt engineering
 - 監督式微調 Supervised finetuning (SFT)
 - 基於強化學習 (RL) 的微調 Reinforcement Learning (RL) based finetuning
 - 自洽性 Self-consistency
 - 檢索-增強-生成 (RAG) 和智能體人工智慧 Retrieve-Augment-Generate (RAG) and Agentic AI
 - OpenClaw Agents

☑ 苦澀的教訓 The Bitter Lesson

□ 世界模型與通往超級智能之路 The World Model and The road to AGI

苦澀的教訓 The Bitter Lesson

By Richard Sutton, 2025 Turing Award Winner, March 13, 2019

國際象棋 (Chess)

早期成功：人類啟發式方法，專家設計的評估函數

最終統治：窮舉搜尋 + 學習 (深藍 → AlphaZero)

圍棋 (Go)

長期以來，人們認為圍棋需要人類的直覺

AlphaGo 的成功主要歸功於首先使用人的數據進行監督訓練，
然後自我對弈和電腦運算

AlphaZero 的成功主要歸功於大規模的自我對弈和電腦運算，
結合強化學習 (RL) 算法 (Unified policy-value network)

語音識別 (Speech Recognition)

手工設計 (Hand-engineered) 的流程被端到端的統計學習取代

電腦視覺 (Computer Vision)

特徵工程 (Man-made Feature Engineering) 被深度學習所取代

苦澀的教訓 The Bitter Lesson

By Richard Sutton, 2025 Turing Award Winner, March 13, 2019

- 不要引入過多阻礙可擴展化 (scalable) 的人類知識。
- 如果LLM完全由人類知識驅動，它將無法有效擴展 (scaling)。

人工智慧領域最大的成功來自於能夠隨著運算能力擴展的通用方法，而不是對人類知識的編碼
—無論我們多麼希望情況並非如此，這個教訓都在不斷重複。

Sutton's OaK (選項與知識 Option and Knowledge) 可擴展架構



- **選項與知識**應該透過計算來學習和改進，而不是用人類的洞察力來編碼。
- 以經驗和獎勵為驅動力進行學習

大綱 (Outline)

□ 大型語言模型及其推理方式

(Large Language Models and How They Reason)

- LLM推理 (LLM reasoning)
 - 思維鏈與提示工程 Chain of thought (CoT) and prompt engineering
 - 監督式微調 Supervised finetuning (SFT)
 - 基於強化學習 (RL) 的微調 Reinforcement Learning (RL) based finetuning
 - 自洽性 Self-consistency
 - 檢索-增強-生成 (RAG) 和智能體人工智慧 Retrieve-Augment-Generate (RAG) and Agentic AI
 - OpenClaw Agents

□ 苦澀的教訓 The Bitter Lesson

☑ 世界模型與通往超級智能之路 The World Model and The road to AGI

世界模型 The World Model

世界模型是潛在空間 (Latent Space) 中學習到的、預測性的、受行動條件影響的模擬器，它能夠實現想像、規劃和數據高效的學習

這與 Sutton 在 OaK 等架構中強調的通用、可擴展的知識以及 Bitter Lesson 強調學習而不是手工編寫的規則密切相關。

生成性 (Generative)：能生成一个自治的三维世界状态。

- 输出不只是语言或图像，而是世界的物理结构、因果关系和演化过程。
- 模型应理解质量、力、速度、摩擦、遮挡、能量守恒等规律。

多模态 (Multimodal)：能整合视觉、听觉、触觉、动作等感知。

- 不同感官信息应映射到同一空间坐标中，从而具备立体理解。
例如：机器人看到一只杯子（视觉），听到水声（听觉），感觉滑动（触觉），都指向同一个对象。

交互性 (Interactive)：能基于反馈持续修正自身世界模型。

- 每个动作的结果（例如抓取成功或失败）都用于更新模型的预测。
- 智能不在于预测静态文本，而在于在动态环境中“活下来”。

“从文字到世界的过渡”。它不是更大的模型，而是更真实的模型。

– Feifei Li, World Labs

- Feifei Li, “From Words to Worlds: Spatial Intelligence is AI’s Next Frontier”, <https://drfeifei.substack.com/p/from-words-to-worlds-spatial-intelligence>
- <https://www.youtube.com/watch?app=desktop&v=s2yLdM46jcM>

世界模型 The World Model

世界模型是潛在空間 (Latent Space) 中學習到的、預測性的、受行動條件影響的模擬器，它能夠實現想像、規劃和數據高效的學習

這與 Sutton 在 OaK 等架構中強調的通用、可擴展的知識以及 Bitter Lesson 強調學習而不是手工編寫的規則密切相關。

生成性 (Generative)：能生成一个自治的三维世界状态。

- 输出不只是语言或图像，而是世界的物理结构、因果关系和演化过程。
- 模型应理解质量、力、速度、摩擦、遮挡、能量守恒等规律。

多模态 (Multimodal)：能整合视觉、听觉、触觉、动作等感知。

- 不同感官信息应映射到同一空间坐标中，从而具备立体理解。
例如：机器人看到一只杯子（视觉），听到水声（听觉），感觉滑动（触觉），都指向同一个对象。

交互性 (Interactive)：能基于反馈持续修正自身世界模型。

- 每个动作的结果（例如抓取成功或失败）都用于更新模型的预测。
- 智能不在于预测静态文本，而在于在动态环境中“活下来”。

“从文字到世界的过渡”。它不是更大的模型，而是更真实的模型。

- Feifei Li, World Labs

- Feifei Li, "From Words to Worlds: Spatial Intelligence is AI's Next Frontier", <https://drfeifei.substack.com/p/from-words-to-worlds-spatial-intelligence>
- <https://www.youtube.com/watch?app=desktop&v=s2yLdM46jcM>

通往超級智能（AGI）之路

- 大規模表徵學習 massive representation learning（基礎模型 Foundation models）
 - LLM模型缺乏事實根基、因果關係和規劃
 - 需要可信的推理、有依據的因果關係和代理人能動性
- 用於預測和規劃的世界模型（World models for prediction and planning）

需要在開放式環境中保持可靠性和保真度，具備穩健的長期自主性
- 透過選項進行分層控制（hierarchical control via options）

需要自動抽象（Automatic abstraction）發現
- 持續學習與記憶（continual learning and memory）

需要持續學習，永不遺忘
- 自主改進（self-directed improvement）

需要穩定、自我演練/自我評估、驗證，並能隨著能力和安全性的提升而擴展。

- Nathan Benaich, Stateof.ai 2025.
- State of AI 2025報告解讀, <https://www.youtube.com/watch?v=oc2T6Nk6Gqw>
- Hendrycks, D., et al., "A Definition of AGI". arXiv preprint arXiv:2510.18212.
- Dawn P. Flanagan, Shauna G. Dixon, "The Cattell-Horn-Carroll Theory of Cognitive Abilities", <https://doi.org/10.1002/9781118660584.ese0431>
- Alexia Jolicoeur-Martineau, "Less is More: Recursive Reasoning with Tiny Networks", Samsung SAIL Montréal, Oct 2025. <https://arxiv.org/html/2510.04871v1>

基督徒从AI的原理中学到什么

羅馬書 12:2

不要效法這個世界，只要心意更新而變化，叫你們察驗何為神的善良、純全、可喜悅的旨意。

1. 学会使用AI工具，有效帮助我们属灵成长；同时谨慎分辨真理与谬误
2. 理解神的话是需要推理过程的，只记住答案不能帮助我们理解与成长。
3. 不经过推理思考，我们的信仰可能不扎实。同时，我们可能也不会用使人信服的智慧言语去传福音。